

Distributed System Providing Scalable Methodology for Real-Time Control of Server Pools and Data Centers

Abstract

A distributed system providing scalable methodology for real-time control of server pools and data centers is described. In one embodiment, a method is described for regulating resource usage by a plurality of programs running on a plurality of machines, the method comprises steps of: providing a resource policy specifying allocation of resources amongst the plurality of programs; determining resources available at the plurality of machines; detecting requests for resources by each of the plurality of programs running on each of the plurality of machines; periodically exchanging resource information amongst the plurality of machines, the resource information including requests for resources and resource availability at each of the plurality of machines; and at each of the plurality of machines, allocating resources to each program based upon the resource policy and the resource information.